

Network Analysis: Overview

Hannu Toivonen

Department of Computer Science and HIIT
P.O. Box 68, FI-00014 University of Helsinki, Finland
`firstname.lastname@cs.helsinki.fi`

Heterogeneous information networks or BisoNets, as they are called in the context of bisociative knowledge discovery, are a flexible and popular form of representing data in numerous fields. Additionally, such networks can be created or derived from other types of information using, e.g., the methods given in Part II of this volume.

This part of the book describes various network algorithms for the exploration and analysis of BisoNets. Their general goal is to support and partially even automate the process of bisociation. More specific goals are to allow navigation of BisoNets by indirect and predicted relationships and by analogy, to produce explanations for discovered relationships, and to help abstract and summarise BisoNets for more effective visualisation.

Contributions

In the first chapter of this part, Dries *et al.* [1] propose BiQL, a novel query language for BisoNets. It is motivated by the observation that graph and network databases have specific needs for query tools, but the tools are much less developed than for relational data. For instance, a statistic such as the shortest path between two given nodes cannot be computed by a relational database. BiQL allows for querying and analyzing databases, especially probabilistic graphs, by using such aggregates and ranking.

The next three chapters address the problem of simplifying a large BisoNet and providing a smaller version instead, both to aid visual exploration and to ease the use of computationally more demanding methods. The first of these chapters, by Zhou *et al.* [2], is an overview of existing approaches to this problem.

The next two chapters then propose novel methods for two specific network abstraction tasks. Zhou *et al.* [3] provide methods for so called network simplification. There, the goal is to remove least important edges, i.e., those that have least effect on the quality of connections between any nodes. In this approach, nodes are left intact. In the chapter on network compression, in turn, Toivonen *et al.* [4] obtain a smaller network by merging nodes that have similar neighbours (or roles) in the network. Such a graph can also be uncompressed to obtain an approximate copy of the original graph. Both of these abstraction methods are designed specifically for BisoNets, paying attention to edge weights and maintaining strengths of (indirect) relations between nodes.

Langohr and Toivonen [5] then introduce a method to identify representative nodes in BisoNets, also motivated by the need to produce different simple views

to large networks. They define a probabilistic similarity measure for nodes, and then apply clustering methods to find groups of nodes. Finally, a representative (the medoid) is output from each cluster, to obtain a sample of nodes that is representative for the whole network.

Kötter and Berthold [6] propose a new approach to extract existing concepts, or to detect missing ones, from a BisoNet by means of concept graph detection. Extracted concepts can then be used to create a higher level representation of the data, while discovered missing concepts might lead to new insights by connecting seemingly unrelated information units.

The final two chapters propose two different approaches to discover similarities or associations — or bisociations — in BisoNets. Thiel and Berthold [7] propose a novel way to find non-trivial structural similarities between nodes in a BisoNet. The basic idea is to compare the neighborhoods of the given nodes, also indirect neighbors. The clue of the method is to do this by comparing the patterns of activation spreading from each of the given nodes.

Finally, Nagel *et al.* [8] address the problem of finding domain bridging associations between otherwise weakly connected domains. They propose a method based purely on structural properties of the connections between entities. It first identifies domains and then assesses interestingness of connections between these domains.

Conclusions

The chapters in this part of the book cover a wide range of methods for bisociation network analysis. Many of the methods are directed to making large BisoNets easier to handle and grasp. Also, a multitude of methods were developed to measure relationships or similarities between entities in BisoNets, and to discover interesting relations or concepts.

Automated discovery of actual, useful bisociations seems to be a very difficult problem. This observation is also supported by the experimental work and applications that are described in Part V of this book. Instead, it is more useful to offer the user tools and mechanisms that help her explore the data, and that facilitate her bisociative processes. Part IV below will continue with even stronger focus on interactive exploration methods for BisoNets.

The applications and evaluations in Part V indicate that the overall bisociative methodology, including network analysis methods as its key components, has potential for helping users make genuine discoveries. At the time of writing, network analysis methods and tools described in this part have already been adopted for regular use by end users.

Open Access. This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Dries, A., Nijssen, S., Raedt, L.D.: BiQL: A Query Language for Analyzing Networks. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*. LNCS (LNAI), vol. 7250, pp. 147–165. Springer, Heidelberg (2012)
2. Zhou, F., Mahler, S., Toivonen, H.: Review of BisoNet Abstraction Techniques. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*. LNCS (LNAI), pp. 166–178. Springer, Heidelberg (2012)
3. Zhou, F., Mahler, S., Toivonen, H.: Simplification of Networks by Edge Pruning. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*. LNCS (LNAI), vol. 7250, pp. 179–198. Springer, Heidelberg (2012)
4. Toivonen, H., Zhou, F., Hartikainen, A., Hinkka, A.: Network Compression by Node and Edge Mergers. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*. LNCS (LNAI), vol. 7250, pp. 199–217. Springer, Heidelberg (2012)
5. Langohr, L., Toivonen, H.: Finding Representative Nodes in Probabilistic Graphs. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*. LNCS (LNAI), vol. 7250, pp. 218–229. Springer, Heidelberg (2012)
6. Kötter, T., Berthold, M.R.: (Missing) Concept Discovery in Heterogeneous Information Networks. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*. LNCS (LNAI), vol. 7250, pp. 230–245. Springer, Heidelberg (2012)
7. Thiel, K., Berthold, M.R.: Node Similarities from Spreading Activation. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*. LNCS (LNAI), vol. 7250, pp. 246–262. Springer, Heidelberg (2012)
8. Nagel, U., Thiel, K., Kötter, T., Piatek, D., Berthold, M.R.: Towards Discovery of Subgraph Bisociations. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*. LNCS (LNAI), vol. 7250, pp. 263–284. Springer, Heidelberg (2012)